# Psychometric validation of the Brief Fatigue Inventory (BFI) in adult X-linked hypophosphatemia (XLH)

A Nixon,[1] A Williams,[2] C Theodore-Oklota[3]

[1]Chilli Consultancy, Salisbury, UK; [2]Kyowa Kirin Services Ltd, Marlow, UK; [3]Ultragenyx Pharmaceutical, Novato, CA, USA

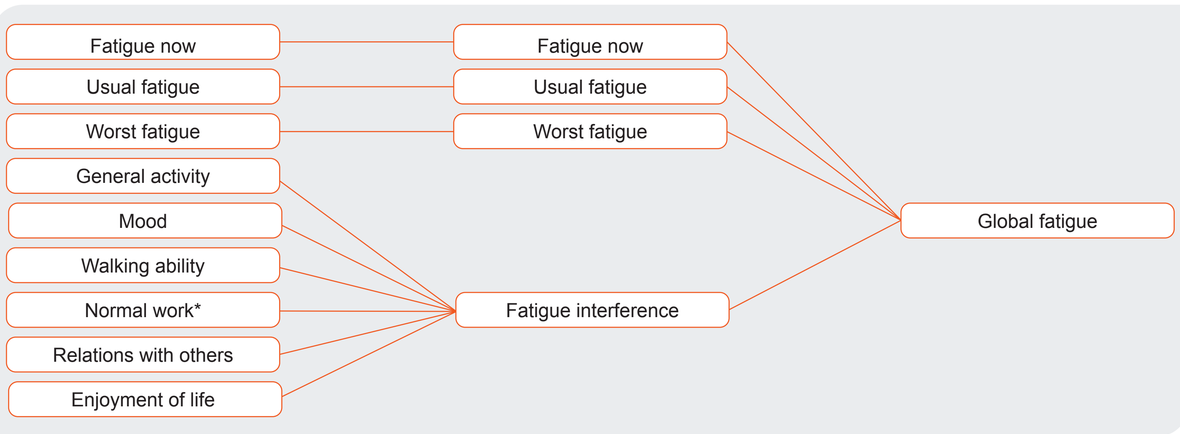
## BACKGROUND

- X-linked hypophosphatemia (XLH) is a rare progressive phosphate-wasting disease. Reduced renal phosphate reabsorption and decreased production of active vitamin D result in chronic hypophosphatemia. Phosphate plays an essential role in cellular energy metabolism and cell signaling[1] as well as bone and teeth mineralization.[2,3]
- XLH typically manifests in childhood as rickets, leading to irreversible skeletal deformity and shortened stature.[2,4] Adults with XLH typically experience pseudofractures and fractures, bone and joint pain, stiffness and fatigue,[4–7] which impair mobility and physical function, compromise quality of life, and limit daily activities.[5,8,9]
- In a qualitative study of 18 adults with XLH, 83% reported fatigue, which many attributed to the impact of XLH on sleep.[9] In another qualitative study of 30 adults with XLH, fatigue was one of the most common symptoms identified by participants, and ranged from mild to severe; most participants reported moderate levels of fatigue.[10]
- The phase 3 UX023-CL303 (NCT02526160) multicenter, randomized, double-blind, placebo-controlled study evaluated the efficacy and safety of burosumab in the treatment of adults with symptomatic XLH. Given the debilitating nature of XLH, several clinical outcome assessments were included to evaluate patients' experience of the disease and its treatment, including the Brief Fatigue Inventory (BFI) to evaluate the impact of treatment on fatigue.
- The current study was undertaken to evaluate the psychometric properties (item/scale properties, reliability, validity, and sensitivity to change) of the BFI in adults with XLH and to determine meaningful change/responder thresholds, using data from the phase 3 study.

## THE BFI

- The BFI was developed to evaluate the severity of fatigue due to cancer and its treatment[11] and has been used in musculoskeletal conditions, including osteoarthritis and rheumatoid arthritis.[12–14] The BFI is a self-administered instrument comprising nine items relating to fatigue. Each is scored on a 0–10 numerical rating scale, with a recall period of 24 hours.[11] Three items measure fatigue, and six measure interference of fatigue on daily life. The items can be combined to form a global score (Figure 1).

***Figure 1. Conceptual framework for the BFI***




*Normal work refers to inside and outside the home.

## OBJECTIVES

- To confirm the item/scale properties, reliability, validity, and sensitivity to change of the BFI in adults with XLH in the context of a phase 3 study
- To establish meaningful change/responder thresholds for Worst Fatigue and Global Fatigue (secondary endpoints in the phase 3 study) and Fatigue Interference (exploratory endpoint)

## METHODS

- Analyses were performed using baseline and week 24 and 48 data from 134 patients in study UX023-CL303 (65% women; age 18–65 years; 66 randomized to burosumab). To be included in the trial, patients had to have bone pain due to XLH/osteomalacia (score ≥4 for Worst Pain on the Brief Pain Inventory [BPI]) at screening.
- The BFI was administered at weeks 0 (baseline), 12, 24, 36, 48, 72, and 96. Patients also recorded responses to the three fatigue severity items (now, usual, worst) in a paper diary for the 7 days preceding visits at weeks 12, 24, 36, and 48.
- The analyses conducted to evaluate the psychometric properties of the BFI are presented in Table 1.

***Table 1. Psychometric evaluation of the BFI***

| Psychometric measure | | Description |
|---|---|---|
| Item response distributions | | Identify skewed distribution and any responses that are over-favored, including floor and ceiling effects |
| Validity | Multi-trait analysis | Evaluates the extent to which each individual item correlates with the domain score it contributes to |
| | Convergent validity | Evaluates correlations of BFI scores with scores on assessments that are conceptually linked |
| | Known groups validity | Evaluates the extent to which the BFI discriminates between groups that are expected to be different |
| Reliability | Internal consistency reliability | Determines the extent to which individual items within the Fatigue Interference and Global Fatigue domains measure the same construct (i.e. homogeneity of the scale) |
| | Test-retest reliability | Reflects the ability of the BFI to give reproducible results when administered twice over a given period to a population with stable disease |
| Responsiveness | | Determines whether observed improvements (or reductions) in scores correspond to improvements (or worsening) in external criteria related to that construct |
| Meaningful change | | Distribution- and anchor-based approaches are used to estimate clinically meaningful change (i.e. minimal clinically important difference; MCID) in the domains of interest |

## RESULTS

### Item response distributions

- The 11 response options (0–10) for each item were mostly well distributed. Item #3 (Worst Fatigue) had the smallest range of responses, with 6 of 11 responses selected, followed by item #2 (Usual Fatigue; 7 of 11). No response options were overly favored for any item.
- Floor effects (>10% respondents endorsing the lowest response on the scale) and ceiling effects (>10% of respondents endorsing the highest response on the scale) were not seen for any item.

### Validity

#### *Item convergent validity*

- All items of the Fatigue Interference and Global Fatigue domains met the item convergent validity criterion (≥0.40), with scale correlations of 0.85–0.92 for Fatigue Interference and 0.72–0.90 for Global Fatigue.

#### *Convergent validity (Table 2)*

- Moderate correlations were seen between Worst Fatigue and five of the seven concurrent validity comparisons: BPI Worst Pain and Pain Interference and Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC®) Pain, Physical Function, and Total Score. Weak correlations were seen for 6-minute walk test (6MWT) and WOMAC Stiffness.
- Moderate or strong correlations were seen between Fatigue Interference and Global Score and six of the seven concurrent validity comparisons. Comparisons were weak for the 6MWT for all three BFI scores.

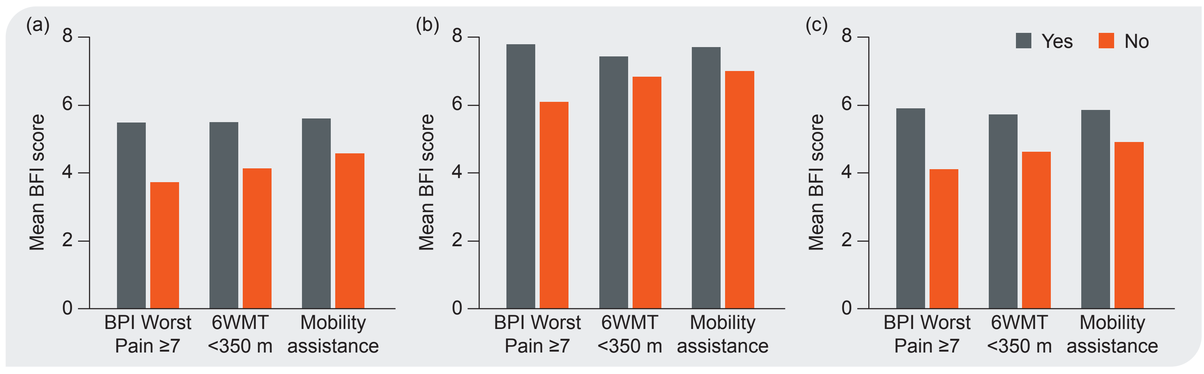
***Table 2. Correlations between the BFI concurrent measures***

| | Worst Fatigue | Fatigue Interference | Global Fatigue |
|---|---|---|---|
| 6MWT (n = 132) | −0.245 | −0.305 | −0.311 |
| BPI Worst Pain (n = 134) | 0.492 | 0.423 | 0.480 |
| BPI Pain Interference (n = 134) | 0.568 | 0.813 | 0.816 |
| WOMAC Pain (n = 133) | 0.515 | 0.562 | 0.603 |
| WOMAC Stiffness (n = 134) | 0.260 | 0.410 | 0.398 |
| WOMAC Physical Function (n = 134) | 0.450 | 0.619 | 0.636 |
| WOMAC Total Score (n = 134) | 0.478 | 0.622 | 0.646 |
| **Key for correlation strength:** Weak <0.39 | moderate, 0.4–0.59 | Strong, ≥0.6 | |

#### *Known groups validity (Figure 2)*

- Known groups analysis was conducted to determine the extent to which scores for BFI Worst Fatigue, Fatigue Interference, and Global Fatigue are able to discriminate groups of subjects who differ in their functioning/health state as defined by other external measures: Worst Pain (BPI-SF 4–6; ≥7); 6MWT distance (<350 m; ≥350 m), and the need for mobility assistance during the 6MWT (yes; no).
- For all three BFI scales, scores for the BPI Worst Pain categories were significantly different between groups (all P<0.001) and 6MWT distance (P=0.033, <0.001, <0.001, for Worst Fatigue, Fatigue Interference, and Global Fatigue, respectively). Scores were not significantly different for patients who needed/did not need mobility assistance during the 6MWT.

***Figure 2. Mean BFI scores for known groups: (a) Fatigue Interference (b) Worst Fatigue (c) Global Fatigue***




### Reliability

#### *Internal consistency reliability*

- **Fatigue Interference:** Cronbach's α coefficient was 0.939 and decreased only slightly when each item was deleted in turn (range 0.919–0.932). This supports the retention of all items and indicates very high internal consistency reliability for this domain. Item total correlations ranged from 0.776 to 0.882.
- **Global Fatigue:** Cronbach's α coefficient was 0.939 and changed minimally when each item was deleted in turn (range 0.926–0.936). This supports the retention of all items and indicates very high internal consistency reliability. Item total correlations ranged from 0.683 to 0.859.

#### *Test-retest reliability (Table 3)*

- Test-retest reliability was evaluated by comparing stability from baseline to week 12 in those endorsing a response of 4 ("no change") on the Patient's Global Impression of Improvement (PGI-I) scale. It was also evaluated by comparing stability from baseline to week 12, week 12 to week 24, and week 24 to week 48 in those with limited improvement in 6MWT (<20 m).[15]
- When the PGI-I was used to define stability, test-retest reliability ranged from 0.677 (Worst Fatigue) to 0.859 (Global Fatigue). Thus, the required benchmark (≥0.70) was reached for Fatigue Interference and Global Fatigue, and was acceptable for Worst Fatigue as a single item scale.
- When the 6MWT was used to define stability, test-retest reliability ranged from 0.433 to 0.616 for Worst Fatigue, 0.676 to 0.783 for Fatigue Interference, and 0.695 to 0.755 for Global Fatigue. Thus, the required benchmark (≥0.70) was reached for Fatigue Interference and Global Fatigue.

***Table 3. Test-retest reliability for the BFI***

| | N | Worst Fatigue | Fatigue Interference | Global Fatigue |
|---|---|---|---|---|
| PGI-I (week 12) | 44 | 0.677 (0.434–0.790) | **0.831 (0.715–0.906)** | **0.859 (0.757–0.921)** |
| 6MWT (baseline–week 12) | 61 | 0.433 (0.085–0.534) | **0.756 (0.576–0.823)** | **0.755 (0.558–0.815)** |
| 6MWT (weeks 12–24) | 61 | 0.559 (0.353–0.707) | **0.783 (0.664–0.865)** | **0.746 (0.615–0.842)** |
| 6MWT (weeks 24–48) | 53 | 0.616 (0.412–0.757) | 0.676 (0.502–0.801) | 0.695 (0.524–0.812) |

Values are intraclass correlations coefficient (95% CI) vs baseline for patients defined as stable
**Bold values** achieved the 0.70 threshold

### Responsiveness

- Changes from baseline to weeks 24 and 48 were compared for groups defined as improved, stable, and worsened using the 6MWT and Patients' Global Impression of Change (PGI-C) (Table 4).

***Table 4. Definitions used in responsiveness analysis***

| | 6MWT definition | PGI-C definition |
|---|---|---|
| Improved | ≥0.5 SD improvement | Responses 1–3 (a little to very much better) |
| Stable | >–0.5 SD to <0.5 SD change | Response 4 (no change) |
| Worsened | ≤0.5 SD worsening | Responses 5–7 (a little to very much worse) |

- **Worst Fatigue:** standardized effect sizes (SES) for the 'improved' group were large for the PGI-C (weeks 24 and 48) and the 6MWT at week 48, and moderate for the 6MWT at week 24.
- **Fatigue Interference:** SES for the 'improved' group were moderate for the PGI-C and 6MWT at week 48, and small for the PGI-C and 6MWT at week 24.
- **Global Fatigue:** SES for the 'improved' group were moderate for the PGI-C (weeks 24 and 48) and the 6MWT at week 48, and small for the 6MWT at week 24.

### MCIDs (Table 5)

***Table 5. Minimal clinically important difference (MCID) estimates and recommended change scores in adults with XLH***

| | Estimated MCID | Change score recommended as MCID in adults with XLH |
|---|---|---|
| Worst Fatigue | −0.80 to −1.73 | −1.5 |
| Fatigue Interference | −0.58 to −1.65 | −1.2 |
| Global Fatigue | −0.50 to −1.59 | −1.2 |

## CONCLUSIONS

- The analyses support the reliability, validity, and responsiveness of the BFI in adults with XLH, and use of this instrument to evaluate the effects of treatment in clinical studies.
- Change scores of –1.5 for BFI Worst Fatigue and –1.2 for the Fatigue Interference and Global Fatigue domains are considered meaningful in adults with XLH.

## REFERENCES

**1.** Muscle fiber contraction and relaxation. In: *Anatomy and Physiology.* 2019, Lumen Learning, Portland OR. **2.** Carpenter TO et al. *Nat Rev Dis Primers* 2017;3:17101. **3.** Carpenter TO et al. *J Bone Miner Res* 2011;26:1381–8. **4.** Linglart A et al. *Endocr Connect* 2014;3:R13–30. **5.** Charles S, Cleeland P. *The Brief Pain Inventory User Guide.* 2009. mdanderson.org/documents/Departments-and-Divisions/Symptom-Research/BPI_UserGuide.pdf. **6.** Skrinar A et al. *J Endocr Soc* 2019;3:1321–34. **7.** Ferizovic N et al. *Adv Ther* 2020;37:770–84. **8.** Hardy DC et al. *Radiology* 1989;171:403–14. **9.** Theodore-Oklota C et al. *Value Health* 2018;21:973–83. **10.** Lo SH et al. *Qual Life Res* 2020, Mar 11 (epub ahead of print). **11.** Mendoza TR et al. *Cancer* 1999;85:1186–96. **12.** Murphy SL et al. *Arthritis Care Res* 2013;65:1617–24. **13.** Murphy SL et al. *Am J Occup Ther* 2010;64:869–76. **14.** Wolfe F. *J Rheumatol* 2004;31:1896–902. **15.** Bennell K et al. *Arthritis Care Res* 2011;63 (suppl 11):S350–70.